# WiFi networks and malware epidemiology

Hao Hu[a,b], Steven Myers[b], Vittoria Colizza[c], and Alessandro Vespignani[b,c,1]

[a]Department of Physics, Indiana University, 727 East Third Street, Bloomington, IN 47405; [b]School of Informatics, Indiana University, 901 East Tenth Street, Bloomington, IN 47408; and [c]Complex Networks Lagrange Laboratory, Institute for Scientific Interchange, 10133 Turin, Italy

Communicated by Giorgio Parisi, University of Rome, Rome, Italy, November 25, 2008 (received for review September 28, 2007)

**In densely populated urban areas WiFi routers form a tightly interconnected proximity network that can be exploited as a substrate for the spreading of malware able to launch massive fraudulent attacks. In this article, we consider several scenarios for the deployment of malware that spreads over the wireless channel of major urban areas in the US. We develop an epidemiological model that takes into consideration prevalent security flaws on these routers. The spread of such a contagion is simulated on real-world data for georeferenced wireless routers. We uncover a major weakness of WiFi networks in that most of the simulated scenarios show tens of thousands of routers infected in as little as 2 weeks, with the majority of the infections occurring in the first 24–48 h. We indicate possible containment and prevention measures and provide computational estimates for the rate of encrypted routers that would stop the spreading of the epidemics by placing the system below the percolation threshold.**

computer security | wireless routers | epidemic spreading

The most common wireless access points are implemented by WiFi routers that supply all of the basic services necessary to access the internet. The use of WiFi routers is becoming close to mainstream in the U.S. and Europe, with 8.4% and 7.9% of all such households having deployed such routers by 2006 (1), and a WiFi market expected to grow quickly in the next few years as more new digital home devices are being shipped with WiFi technology.

As WiFi deployment becomes more and more pervasive, however, there is a larger risk that massive attacks exploiting the WiFi security weaknesses could affect large numbers of users.

Malware is the name given to a broad range of software, including viruses and worms, that has malicious or fraudulent intent. Recent years have witnessed a change in both the designers of malware attacks and their motivations. Malware creators have shifted from programmer enthusiasts attempting to get peer credit from the "hacker" community to organized crime engaging in fraud and money laundering through varying forms of online crime. In this context, WiFi routers represent valuable targets when compared with the PCs that malware traditionally infects, because they are the perfect platform to launch a number of attacks (2–5) that previous security technologies have reasonably assumed were unlikely (6). Unlike PCs, they tend to be always on and connected to the internet, and currently there is no software aimed at specifically detecting or preventing their infection. Routers need to be within relatively close proximity to each other to communicate wirelessly, but an attack can now take advantage of the increasing density of WiFi routers in urban areas that creates large ad hoc geographical networks where the malware can propagate undisturbed, making WiFi vulnerabilities considerably more risky than previously believed (5, 7).

Here, we assess the vulnerability of WiFi networks of different U.S. cities by simulating the wireless propagation of a malicious worm spreading directly from wireless router to wireless router. We construct an epidemiological model that takes into account several widely known and prevalent weaknesses in commonly deployed WiFi routers' security (2, 8) [e.g., default and poor password selection and cracks in the wired equivalent privacy (WEP) cryptographic protocol (9)]. The WiFi proximity networks over which the attack is simulated are obtained from real-world geographic location data for wireless routers. The infection scenarios obtained for a variety of U.S. urban areas are troublesome in that the infection of a small number of routers in most of these cities can lead to the infection of tens of thousands of routers in a week, with most of the infection occurring in the first 24 h. We address quantitatively the behavior of the spreading process, and we provide specific suggestions of increased usage of the WiFi Protected Access (WPA) encryption protocol and strong administrative passwords to minimize the WiFi network weakness and mitigate an eventual attack.

## Results and Discussion

**WiFi Networks.** WiFi routers, even if generally deployed without a global organizing principle, define a self-organized proximity communication network. Indeed, any 2 routers that are in the range of each other's WiFi signal can exchange information and may define an ad hoc communication network. These networks belong to the class of spatial or geometric networks in that nodes are embedded in a metric space, and the interaction between 2 nodes strongly depends on the range of their spatial interaction (10–13).

In this perspective, one might wonder whether the actual deployment of WiFi routers is sufficient at the moment to generate large connected networks spanning sizeable geographic areas. This problem, equivalent to the percolation of giant connected components in graph theory (14, 15), is, however, constrained by the urban area's topology and demographic distribution dictating the geographical locations of WiFi routers. Here, we consider WiFi networks as obtained from the public worldwide database of the Wireless Geographic Logging Engine (WiGLE) website.* The database collects data on the worldwide geographic location of wireless routers and at the time of our study counted >10 million unique WiFi data points on just <600 million observations, providing good coverage of the wireless networks in the U.S. and in North Central Europe. The data provide a wealth of information that include, among other things, the routers' geographic locations [expressed in latitude ($LAT$) and longitude ($LON$)] and their encryption statuses. In particular, we focused on the wireless data extracted from 7 urban areas or regions within the U.S.—Chicago, Boston, New York City, San Francisco Bay Area, Seattle, and Northern and Southern Indiana. Starting from the set of vertices corresponding to georeferenced routers in a given region, we construct the proximity network (10–13) by drawing an edge between any 2 routers $i$ and $j$ located at $\vec{p}_i = (LON_i, LAT_i)$ and $\vec{p}_j = (LON_j, LAT_j)$, respectively, whose geographical distance $d(\vec{p}_i, \vec{p}_j)$ is smaller than the maximum interaction radius $R_{\text{int}}$ (i.e., $d(\vec{p}_i, \vec{p}_j) \leq R_{\text{int}}$). In the WiFi networks, the maximum interaction radius $R_{\text{int}}$

Author contributions: H.H., S.M., V.C., and A.V. designed research, performed research, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

[1]To whom correspondence should be addressed. E-mail: alexv@indiana.edu.

*www.wigle.net.

This article contains supporting information online at www.pnas.org/cgi/content/full/0811973106/DCSupplemental.

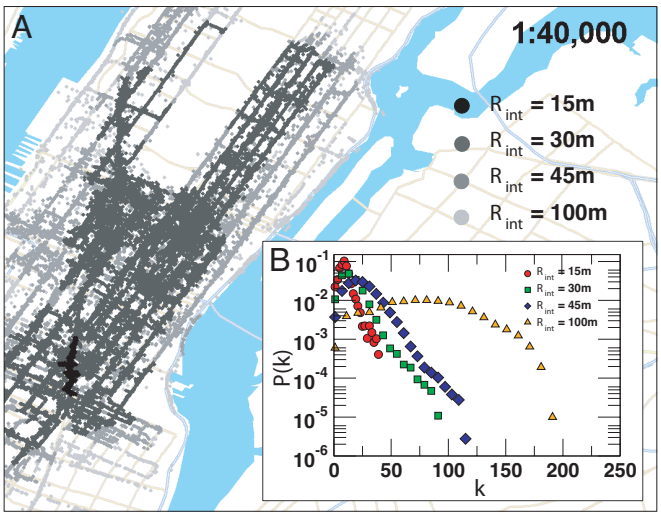



**Fig. 1.** Visualization and degree distribution of the WiFi proximity networks. (*A*) Map representation of the giant components of the WiFi network in the Manhattan area as obtained with different values of $R_{int}$. (*B*) The degree distribution for different values of the interaction radius $R_{int}$ show an exponential decay and a cutoff that depends on $R_{int}$. The result is obtained as averages over 5 different randomization procedures to redefine the location of each router.

strongly depends on the local environment of any specific router. In practice, $R_{int}$ ranges from 15 m for a closed office with poor transmission to ≈100 m outdoors (16). For simplicity, we assume that $R_{int}$ is constant, independent of the actual location of a given router, and we consider 4 different values of the maximum interaction radius—$R_{int} \in${15 m, 30 m, 45 m,100 m}—analyzing the resulting networks for each of the 7 regions under study. A more detailed account of the network construction procedure and the filtering methods used to minimize potential biases introduced by the data collection mechanisms are described in *Materials and Methods*.

In Fig. 1*A*, we report an illustration of the giant component of the network obtained in the Manhattan area for different values of $R_{int}$. It is possible to observe that despite the clear geographical embedding and the city constraints, a large network of >36,000 routers spans the downtown area for $R_{int}$ set to 45 m. The degree distributions of the giant components, i.e., the probability that any giver router is within range and connected to $k$ other routers (see Fig. 1*B*), are characterized by an exponential decrease (12) with a cutoff clearly increasing with the interaction radius, because a larger range increases the number $k$ of nodes found within the signal area. Very similar properties are observed in all of the networks analyzed, and a detailed account of their topology is reported in supporting information (SI). It is important to stress that the metric space embedding exerts a strong preventative force on the small-world behavior of the WiFi networks, because the limited WiFi interaction rules out the possibility of long-range connections.

**Infecting a Router.** The infection of a susceptible router occurs when the malware of an already infected router is able to interface with the susceptible's administrative interface over the wireless channel. Two main technologies aim at preventing such infection through (*i*) the use of encrypted and authenticated wireless channel communication through the WEP and WPA cryptographic protocols and (*ii*) the use of a standard password for access control. Encryption should provide an initial level of security, because it needs to be bypassed before a potential attacker could attempt to bypass the administrative password. Most users do not currently employ their routers' encryption capabilities—indeed the encryption rates in the considered cities vary from 21% to 40% of the population, as shown in *Materials and Methods*. For the purposes of this work, we assume that WPA encryption is not vulnerable to attack, and therefore, any router that uses it is considered immune to the worm. Because of cryptographic flaws in WEP, this protocol can always be broken, given that the attacker has access to enough encrypted communication. This can be achieved by waiting for the router to be used by legitimate clients or by deploying more advanced active attacks. Bypassing WEP encryption is therefore feasible and only requires a given amount of time.

Once the malware has bypassed any cryptographic protocol and established a communication channel, it may then attempt to bypass the password. A large percentage of users do not change their password from the default established by the router manufacturer, and these passwords are easily obtainable. Here, we use as a proxy for this percentage the fraction of users who do not change their routers default service set identifier (SSID). For all of the other routers, we assume that 25% of them can have the password guessed with 65,000 login attempts, based on the evidence provided by security studies (17) that showed that ≈25% of all users' passwords are contained in a dictionary of 65,000 words. We then assume, based on previous worms, that another 11% of passwords are contained in a larger library of approximately a million words (18). No back-off mechanism exists on the routers, which prevents systematic dictionary attacks. In case the password is not found in either dictionary, the attack cannot proceed. Alternatively, if the password has been overcome, the attacker can upload the worm's code into the router's firmware, a process that typically takes just a few minutes. In *Materials and Methods*, we report a list of the typical time scales related to each step of the attack strategy.

**Construction of the Epidemic Model.** The construction of the wireless router network defines the population and the related connectivity pattern over which the epidemic will spread. To describe the dynamical evolution of the epidemic (i.e., the number of infected routers in the population as a function of time), we use a framework analogous to epidemic modeling that assumes that each individual (i.e., each router) in the population is in a given class depending on the stage of the infection (19). Generally, the basic modeling approaches consider 3 classes of individuals: susceptible (those who can contract the infection), infectious (those who contracted the infection and are contagious), and recovered (those who recovered or are immune from the disease and cannot be infected). Analogous schemes have been used in the past to simulate computer viruses spreading on the wired internet and e-mail networks (20–22). These studies have pointed out the importance of the heterogeneity of the internet networks that might eventually lead to the virtual lack of epidemic threshold. The regular internet virus spreads, however, on network topologies where connections and transmissions do not have a finite range, whereas in the present case, the WiFi underlying network is deeply influenced by geographical embedding and by finite range of transmission.

Furthermore, the heterogeneity of the WiFi router population in terms of security attributes calls for an enlarged scheme that takes into account the differences in the users' security settings (other sources of heterogeneity in the router platforms are discussed in SI). We consider 3 basic levels of security and identify the corresponding classes: routers with no encryption, which are potentially the most exposed to the attack, are mapped into a first type of susceptible class $S$; routers with WEP encryption, which provides a certain level of protection that can be eventually overcome with enough time, are mapped into a second type of susceptible class denoted $S_{WEP}$; routers with WPA encryption, which are assumed to resist any type of attacks, correspond to the removed class $R$. This classification, however,

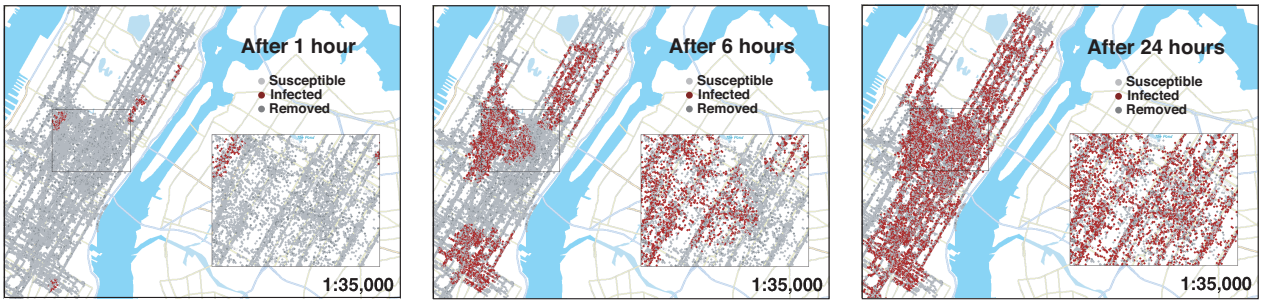



**Fig. 2.** Illustration of the spread of a wireless worm through Manhattan in several time slices. In this series, the result is based on 1 randomization procedure for the location of each router and the maximum interaction radius $R_{\text{int}}$ is set to 45 m.

needs to be refined to take into account the password settings of the users that range from a default password to weak or strong passwords and finally to noncrackable passwords. For this reason, we can think of the nonencrypted class $S$ as being subdivided into 4 subclasses. First, we distinguish between the routers with default password $S_{\text{nopass}}$ and the ones with a password $S_{\text{pass1}}$. The latter contains routers with all sorts of passwords that undergo the first stage of the attack that employs the smaller dictionary. If this strategy fails, the routers are then classified as $S_{\text{pass2}}$ and undergo the attack that employs the larger dictionary. Finally, if the password is unbreakable, the router is classified as $R_{\text{hidden}}$. The last class represents routers whose password cannot be bypassed. However, their immune condition is hidden in that it is known only to the attacker who failed in the attempt, whereas for all of the others, the router appears in the susceptible class as it was in its original state. This allows us to model the unsuccessful attack attempts of other routers in the dynamics. WEP encrypted routers have the same properties in terms of password, but the password relevance starts only when the WEP encryption (if any) has been broken on the router. At this stage of the attack it can be considered to be in the nonencrypted state, and therefore no subclasses of $S_{\text{WEP}}$ have to be defined. In addition to the above classes, the model includes the infected class ($I$) with those routers that have been infected by the malware and have the ability to spread it to other routers.

The model dynamics is specified by the transition rates among different classes for routers under attack. Transitions will occur only if a router is attacked and can be described as a reaction process. For instance the infection of a nonencrypted router with no password is represented by the process $S_{\text{nopass}} + I \rightarrow 2I$. The transition rates are all expressed as the inverse of the average time needed to complete the attack. In the above case, the average time of the infection process is $\tau = 5$ min and the corresponding rate $\beta$ for the transition $S_{\text{nopass}} + I \rightarrow 2I$ is $\beta = \tau^{-1}$. Similarly the time scale $\tau_{\text{WEP}}$ needed to break a WEP encryption will define the rate $\beta_{\text{WEP}}$ ruling the transition from the $S_{\text{WEP}}$ to the nonencrypted class. In *Materials and Methods*, we report in detail all of the transition processes and the associated rates defining the epidemic processes.

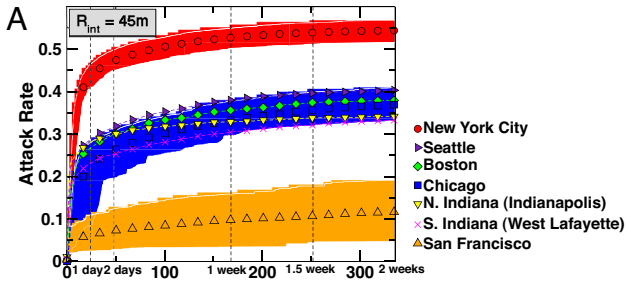
One of the most common approaches to the study of epidemic processes is to use deterministic differential equations based on the assumption that individuals mix homogeneously in the population, each of them potentially in contact with every other (19). In our case, the static nonmobile nature of wireless routers and their geographical embedding make this assumption completely inadequate, showing the need to study the epidemic dynamics by explicitly considering the underlying contact pattern (21, 23–26). For this reason, we rely on numerical simulations obtained by using an individual-based modeling strategy. At each time step, the stochastic disease dynamics are applied to each router by considering the actual state of the router and those of its neighbors as defined by the actual connectivity pattern of the network. It is then possible to measure the evolution of the number of infected individuals and keep track of the epidemic progression at the level of single routers. In addition, given the stochastic nature of the model, different initial conditions and stochastic noise realizations can be used to obtain different evolution scenarios.

Because multiple-seed attacks are likely, we report simulations with initial conditions set with 5 infected routers randomly distributed within the population under study. Single-seed attacks and different number of initial seeds have similar effects and are reported in SI. The initial state of each router is directly given by the real WiFi data or is obtained from estimates based on real data, as detailed in *Materials and Methods*. Finally, for each scenario, we report the averages of >100 realizations. Reports on single realizations and their properties are in SI.

**Spreading of Synthetic Epidemics.** According to the simulation procedure outlined above, we study the behavior of synthetic epidemics in the 7 urban areas we used to characterize the properties of WiFi router networks. The urban areas considered are quite diverse in that they range from a relatively small college town (West Lafayette, IN) to big metropolises such as New York City and Chicago. In each urban area, we focus on the giant component of the network obtained with a given $R_{\text{int}}$ that may vary consistently in size.

Here, we report the results for a typical epidemic spreading scenario in which the time scales of the processes are chosen according to their average estimates. In the SI, we report the best- and worst-case scenarios obtained by considering the combination of parameters that maximize and minimize the rate of success of each attack process, respectively. The networks used as substrate are obtained in the intermediate interaction range of 45 m. The sensitivity analysis to the change of this parameter is reported in SI.

The 3 snapshots of Fig. 2 provide an illustration of the evolution of a synthetic epidemic in the Manhattan area; shown in red are the routers that are progressively infected by malware. The striking observation is that the malware rapidly propagates on the WiFi network in the first few hours, taking control of ≈55% of the routers after 2 weeks from the infection of the first router. The quantitative evidence of the potential impact of the epidemic is reported in Fig. 3*A*, where the average profile of the density of infected routers is reported for all of the urban areas considered in the numerical experiment. Although it is possible to notice a considerable difference among the various urban areas, in all cases, we observe a sharp rise of the epidemic within the first couple of days and then a slower increase, which after 2 weeks leaves ≈10% to 55% of the routers in the giant component controlled by malware. The similar time scale in the rise of the epidemic in different urban areas is not surprising because it is mainly determined by the time scale of the specific attacks considered in the malware spreading model. In general




B

| City | Snopass% | Spass1% | Senc% |
|---|---|---|---|
| New York City | **87.84%** | **31.82%** | **31.05%** |
| Seattle | **66.82%** | **24.58%** | **22.55%** |
| Boston | **60.12%** | **21.89%** | **18.11%** |
| Chicago | **61.53%** | **22.20%** | **19.42%** |
| S.Indiana | **50.93%** | **18.07%** | **17.27%** |
| N.Indiana | **52.59%** | **19.02%** | **16.18%** |
| San Francisco | **19.63%** | **6.95%** | **6.04%** |

**Fig. 3.** Impact of the epidemic. (*A*) Average attack rate (density of infected routers) versus time for the giant component of all of the 7 urban areas, and 90% C.I. for 3 prototypical cases, keeping $R_{int}$ = 45 m. (*B*) Fraction of infected routers in classes with different security level.

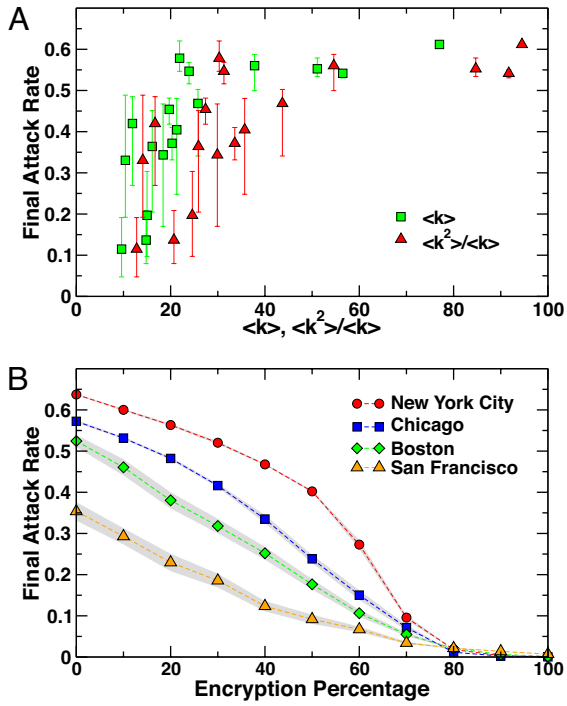



**Fig. 4.** Impact of topology and encryption usage on the epidemic. (*A*) The correlation between the final attack rate and average degree as well as degree fluctuations. (*B*) Attack rate as a function of a differing fraction of encrypted routers in 4 different urban areas. A larger fraction of encrypted routers drastically reduces the impact of the epidemics. When the fraction of encrypted routers falls between 60% and 70%, the urban areas exhibit a network that is below the percolation threshold for the epidemic, and the attack rate is close to zero.

the sharp rise of the epidemic in its early stages is due to the nonencrypted routers that are infected in a very short time. This is clearly shown in Fig. 3*B*, where the fraction of infected routers belonging to different classes is reported. Obviously, nonencrypted routers are those that are most affected by the epidemic. The slower progression at later stages is instead due to the progressive infection of WEP routers whose attack time scale is ≈1 order of magnitude longer (see the SI for more details on single-realizations behavior).

A more complicated issue is understanding the different attack (infection) rates that the epidemic attains in different urban area networks. The pervasiveness of the epidemic can be seen as a percolation effect on the WiFi network (27, 28). The WPA-encrypted routers and those with unbreakable passwords represent obstacles to the percolation process and define an effective probability that each router may be infected at the end of the spreading process. This probability has to be compared with the percolation probability threshold of the network, above which it is possible to have a macroscopic spanning cluster of connected and infected routers (28–30). The larger the effective percolation probability with respect to the threshold, the larger the final density of infected routers. On the other hand, the epidemic thresholds of the networks are not easy to estimate because they are embedded in the particular geometries of the cities' geographies. In random networks, large average degree and large degree fluctuations favor the spreading of epidemics and tend to reduce the network percolation threshold (21, 31). Fig. 4*A* shows an appreciable statistical correlation between the attack rate and these quantities. On the other hand, there are many other network features that affect the percolation properties of the networks. First, the cities have different fractions of encrypted routers. Although these fractions are not extremely dissimilar, it is clear that, given the nonlinear effect close to the percolation threshold, small differences may lead to large difference in the final attack rate. For instance, San Francisco, with the largest fraction of encrypted routers corresponding to ≈40% of the population, exhibits the smallest attack rate among all of the urban areas considered. Second, the geometrical constraints imposed by the urban area geography may have a large impact on the percolation threshold, which can be rather sensitive to the local graph topology. For instance, network layouts with 1D bottlenecks or locally very sparse connectivity may consistently lower the attack rate by sealing part of the network, and thus protecting it from the epidemic. Indeed, a few WPA routers at key bottlenecks can make entire subnetworks of the giant component impenetrable to the malware.

## Conclusions

Based on the previous results, we note that there is a real concern about the wireless spread of WiFi-based malware. This suggests that action needs to be taken to detect and prevent such outbreaks, and more thoughtful planning for the security of future wireless devices needs to occur, so that such scenarios do not occur or worsen with future technology. For instance, given the increasing popularity of the IEEE 802.11n standard for WiFi networks with its increased wireless communications range, the possibility for larger infections to occur is heightened, because of the larger connected components that will emerge (see SI). Furthermore, it is highly likely that we will only see the proliferation of more wireless standards as time goes by, and all of these standards should consider the possibility of such epidemics.

There are 2 preventive actions that can be easily considered to successfully reduce the rates of infection. First, force users to change default passwords, and second, the adoption of WPA, the cryptographic protocol meant to replace WEP, which does not share its cryptographic weaknesses. In Fig. 4*B*, we report the impact of the epidemic when we progressively increase the fraction of routers with encryption. We perform the experiment under the restrictive assumption that, among the encrypted routers, only the usual 30% uses the safe WPA and keep the same statistics for the password choices as for the baseline simulations. The fraction of infected routers after 2 weeks is quickly dropping when the encryption percentage falls between 60% and 70%. This would correspond to

a fraction of immune WPA routers of ≈20% to 30% at which the percolation threshold is reached, and the epidemic is not able to spread across the network. It should be noted that because of the different topologies of the networks in different cities, we should not expect a single percolation threshold to hold for all locals. In SI, we provide a more precise measurement of the threshold by using as a proxy the divergence of the average infected cluster size (28, 32). Finally, better results can be achieved by improving the password choices in the rest of WEP-encrypted routers.

Unfortunately, the dangers of poorly chosen user passwords have been widely publicized for at least 2 decades now, and there has been little evidence of a change in the public's behavior. In addition, there are many barriers to public adoption of WPA on wireless routers. The use of only 1 device in the home that does not support WPA, but that does support the more widely implemented WEP, is sufficient to encourage people to use WEP at home. For this reason, a detailed study of the impact of targeted deployment of WPA routers in key locations of the network needs to take place.

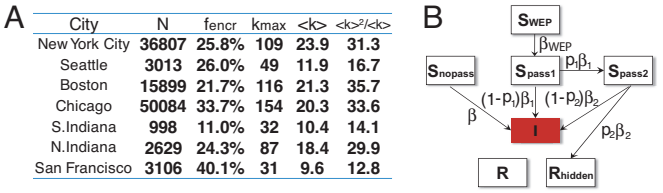
## Materials and Methods

**WiFi Data and Networks.** WiFi data are downloaded from wigle.net for 7 urban areas in the U.S. and is processed to eliminate potential biases introduced by data collection. Records that appear as probe in their type classification are removed from the dataset because they correspond to wireless signals originating from nonrouters. Such records represent a very small percentage of the total number in every city considered. For example, in the urban area of New York City, there were 2,586 probe records, corresponding to 5.4% of the total (additional details for all urban areas under study are provided in SI).

A preliminary spatial analysis of the data for each urban area reveals the presence of sets of several WiFi routers sharing an identical geographic location. To avoid biases due to overrepresentation of records, we checked for unique basic service set identifier (BSSID) (i.e., MAC address) and assume that each of these locations could contain at most $n$ overlapping routers, where $n$ was fixed at 20 to provide a realistic scenario, such as a building with several hot spots. For New York City, this procedure led to the elimination of 216 records, which represent 0.5% of the total number of WiFi routers. This also takes into account the exclusion of the vertical dimension of the problem, namely the presence of WiFi routers in very tall buildings, that is, however, not relevant for the geographical spreading of the epidemics.

More importantly, we adopt a randomized procedure to redefine the position of each router in a circle of radius $R_{ran}$ centered on the GPS coordinates provided by the original data. This procedure is applied to approximate the actual location of each router, which would be otherwise localized along city streets, due to an artifact of the wardriving data collection method. The newly randomized positions for the set of routers completely determine the connectivity pattern of the spatial WiFi network and its giant component substrate for the epidemic simulation. Results presented here are obtained as 5 averages over several randomization procedures. Fig. 5*A* reports the main topological indicators of the giant components of each urban area extracted from the WiFi network built assuming that $R_{int} = 45$ m. It is important to stress that the many properties cannot be easily deduced by models based on uniform distribution of points in a 2D Euclidean space, because the emerging degree and clustering distribution are deeply affected by the geographical and demographic properties of each given urban area.

**Epidemic Model.** Fig. 5*B* shows the flow diagram of the transmission model. Initial conditions set the number of routers belonging to each of the following compartments: $S_{nopass}$ (routers with no encryption and default password), $S_{pass1}$ (routers with no encryption and user set password), $S_{WEP}$ (routers with WEP encryption), and $R$ (routers with WPA encryption, here considered immune). The classes $S_{pass2}$ and $R_{hidden}$ are void at the beginning of the simulations because they represent subsequent stages of the infection dynamics. Encrypted routers are identified from original data, and the fraction of $R$ of the total number of encrypted routers is assumed to be 30%, in agreement with estimates on real-world WPA usage. Analogously, we assume that the nonencrypted routers are distributed according to the following proportions: 50% in class $S_{nopass}$ and 50% in class $S_{pass1}$.



A

| City | N | fencr | kmax | <k> | <k²>/<k> |
|---|---|---|---|---|---|
| New York City | 36807 | 25.8% | 109 | 23.9 | 31.3 |
| Seattle | 3013 | 26.0% | 49 | 11.9 | 16.7 |
| Boston | 15899 | 21.7% | 116 | 21.3 | 35.7 |
| Chicago | 50084 | 33.7% | 154 | 20.3 | 33.6 |
| S.Indiana | 998 | 11.0% | 32 | 10.4 | 14.1 |
| N.Indiana | 2629 | 24.3% | 87 | 18.4 | 29.9 |
| San Francisco | 3106 | 40.1% | 31 | 9.6 | 12.8 |



**Fig. 5.** WiFi networks' properties and epidemic transmission model. (*A*) Properties of the WiFi network giant components for $R_{int} = 45$ m: size of the giant component $N$; percentage of encrypted routers, $f_{encr}$; maximum degree, $k_{max}$; average degree, $\langle k \rangle$; degree fluctuations, $\langle k^2 \rangle / \langle k \rangle$. The results presented are obtained as averages over 5 different randomization procedures to redefine the location of each router. (*B*) Compartmental flows for the epidemic model.

The infection dynamics proceeds as follows. A router with no encryption enters the infectious class with unitary rate if attacked. The attack to a router in class $S_{pass1}$ is characterized by a transition rate $\beta_1$ and has 2 possible outcomes: with probability $(1 - p_1)$, the router is infected and enters $I$, whereas with probability $p_1$ it enters $S_{pass2}$ because the attacker is not able to overcome the password, and the infection attempt requires additional time and resources. Once in class $S_{pass2}$, it can become infectious with probability $(1 - p_2)$ if the attack is successful, or otherwise the router enters $R_{hidden}$ with probability $p_2$ because the password has not been bypassed. This process occurs with a transition rate $p_2$. WEP-encrypted routers follow the same dynamics once the encryption is broken, and they enter $S_{pass1}$ with transition rate $\beta_{WEP}$ .We do not allow the transition between $S_{WEP}$ and $S_{nopass}$ because we assume that anyone who went to the trouble of enabling encryption would also go to the trouble of changing the default password.

The numerical simulations consider the discrete nature of the individuals and progress in discrete time steps. We assume that the attacker will target the router, among its neighbors, with the lowest visible security settings. In addition, we do not allow simultaneous attacks, so that each infected router will choose its next target only among those routers that are not already under attack. Once an attack has started, the attacker will keep trying to bypass the security setting of the same target until the attempt is finally successful or not. In both cases, the attacker will then move to another target. The simulation's unitary time step is defined by the shortest time scale among all processes involved, i.e., the time $\tau$ needed to complete an attack to a nonencrypted router with no password. This automatically defines as unitary the transition rate associated to the reaction $S_{nopass} + I \rightarrow 2I$. Typical time scales for the other processes are: $\tau_1 = 6$–$15$ min to bypass a password in the smaller dictionary, $\tau_2 = 400$–$1{,}000$ min to bypass a password in the larger dictionary, $\tau_{WEP} = 2{,}880$–$5{,}760$ min to crack the WEP encryption. The corresponding transition rates can be analogously defined as probabilities expressed in terms of their ratio with $\beta$ that defines the unitary rate.

Simulations run for 4,032 time steps, corresponding to 20,160 min (i.e., 2 weeks). At each time step, we measure the global attack rate defined as the number of infectious $I(t)$ at time $t$ over the total population of the network discounted by the number of recovered, $N - R$. In this way, we can take into account the differences of the encryption percentages observed in different urban areas.

**ACKNOWLEDGMENTS.** H.H. thanks the Institute for Scientific Interchange in Turin for its hospitality during the time this work was completed. A.V. was partially supported by National Science Foundation Grant IIS-0513650 and National Institutes of Health Grant R21- DA024259.

1. Mercer D (2006) *Home Network Adoption: Wi-Fi Emerges as Mass Market Phenomenon.* (Market Report, Strategy Analytics, Newton, MA).
2. Stamm S, Ramzan Z, Jakobsson M (2006) *Drive-by Pharming.* (Tech Rep 641, Indiana Univ, Bloomington, IN).
3. Ollmann G (2006) *The Pharming Guide.* (Tech Rep, Next Generation Security Software Ltd., Sutton, UK).
4. Jakobsson M, Myers S, eds (2007) *Phishing and Countermeasures: Undertanding the Increasing Problem of Electronic Identity Theft.* (Wiley, New York).
5. Akritidis P, Chin WY, Lam VT, Sidiroglou S, Anagnostakis KG (2007) Proximity breeds danger: Emerging threads in metro-area wireless networks. *Proc 16th USENIX Security Symposium* (USENIX, Berkeley, CA), pp 323–338.
6. Myers S, Stamm S (2008) Practice and prevention of home-router mid-stream injection attacks. *IEEE Proc Anti-Phishing Working Group eCrime Research Summit, 2008*, IEEE, Washington, DC), in press.
7. Traynor P, Butler K, Enck W, Borders K, McDaniel P (2006) *Malnets: Large-Scale Malicious Networks via Compromised Wireless Access Points.* (Tech Rep NAS-TR-0048-2006, Network and Security Research Center, Pennsylvania State Univ, State College, PA).
8. Tsow A, Jakobsson M, Yang L, Wetzel S (2006) Warkitting: The drive-by subversion of wireless home routers. *J Digital Forensic Practice* 1:179–192.
9. Bittau A, Handley M, Lackey J (2006) The final nail in WEP's coffin. *SP '06: Proceedings of the 2006 IEEE Symposium on Security and Privacy* (IEEE Comput Soc Washington, DC), pp 386–400.

10. Dall J, Christensen M (2002) Random geometric graphs. *Phys Rev E* 66:016121–016130.
11. Nemeth G, Vattay G (2003) Giant clusters in random ad hoc networks. *Phys Rev E* 67:036110–036116.
12. Herrmann C, Barthélemy M, Provero P (2003) Connectivity distribution of spatial networks. *Phys Rev E* 68:026128–026134.
13. Helmy A (2003) Small worlds in wireless networks. *IEEE Comm Lett* 7:490–492.
14. Molloy M, Reed B (1995) A critical point for random graphs with a given degree sequence. *Random Struct Algorithm* 6:161–179.
15. Bollobas B, Riordan O (2006) *Percolation* (Cambridge Univ Press, Cambridge, UK).
16. Gast M (2005) *802.11 Wireless Networks: The Definitive Guide* (O'Reilly, Sebastopol, CA), Second Ed.
17. Klein DV (1990) Foiling the Cracker: A Survey of, and Improvements to, Password Security. *Proc Second USENIX Workshop on Security* (USENIX, Berkeley, CA), pp 5–14.
18. Jeff J, Alan Y, Ross B, Alasdair A (2000) *The Memorability and Security of Passwords—Some Empirical Results.* (Tech Rep No. 500, Computer Laboratory, Univ of Cambridge, Cambridge, UK).
19. Anderson RM, May RM (1992) *Infectious Diseases of Humans: Dynamics and Control* (Oxford Univ Press, Oxford).
20. Kephart JO, White SR (1993) Measuring and modeling computer virus prevalence. *Proc 1993 IEEE Comput Soc Symp on Res in Security and Privacy* (IEEE, Washington, DC).
21. Pastor-Satorras R, Vespignani A (2001) Epidemic spreading in scale-free networks. *Phys Rev Lett* 86:3200–3203.
22. Balthrop J, Forrest S, Newman MEJ, Williamson MM (2004) Technological networks and the spread of computer viruses. *Science* 304:527–529.
23. Watts DJ, Strogatz SH (1998) Collective dynamics of ''small-world'' networks. *Nature* 393:440–442.
24. Barabási AL, Albert R (1999) Emergence of scaling in random networks. *Science* 286:509–512.
25. Keeling MJ (1999) The effects of local spatial structure on epidemiological invasions. *Proc R Soc London Ser B* 266:859–867.
26. Moore C, Newman MEJ (2000) Epidemics and percolation in small-world networks. *Phys Rev E* 61:5678–5682.
27. Grassberger P (1983) Critical behavior of the general epidemic process and dynamical percolation. *Math Biosci* 63:157.
28. Ben-Avraham B, Havlin S (2000) *Diffusion and Reactions in Fractals and Disordered Systems* (Cambridge Univ Press, Cambridge, UK).
29. Cohen R, Erez K, Ben-Avraham D, Havlin S (2000) Resilience of the internet to random breakdowns. *Phys Rev Lett* 85:4626–4628.
30. Callaway DS, Newman MEJ, Strogatz SH, Watts DJ (2000) Network robustness and fragility: Percolation on random graphs. *Phys Rev Lett* 85:5468–5471.
31. Lloyd AL, May RM (2001) How viruses spread among computers and people. *Science* 292:1316–1317.
32. Carmi S, Havlin S, Kirkpatrick S, Shavitt Y, Shir E (2007) A model of internet topology using k-shell decomposition. *Proc Natl Acad Sci USA* 104:11150–11154.